

J. Tung

R-run



1637

## RAW SEQUENCE LISTING

DATE: 02/28/2002

PATENT APPLICATION: US/09/891,332A

TIME: 15:05:17

Input Set : N:\Crf3\Refhold\I891332A.raw

Output Set: N:\CRF3\02282002\I891332A.raw

## SEQUENCE LISTING

## 1 (1) GENERAL INFORMATION:

2 (i) APPLICANT: Chatterjee, Deb K.

3 Solus, Joseph

4 Yang, Shuwei

5 (ii) TITLE OF INVENTION: Polymerases for Analyzing or Typing Polymorphic

6 Nucleic Acid Fragments and Uses Thereof

7 (iii) NUMBER OF SEQUENCES: 93

8 (iv) CORRESPONDENCE ADDRESS:

9 (A) ADDRESSEE: STERNE, KESSLER, GOLDSTEIN &amp; FOX, P.L.L.C

10 (B) STREET: 1100 New York Ave., N.W., Suite 600

11 (C) CITY: Washington

12 (D) STATE: DC

13 (E) COUNTRY: USA

14 (F) ZIP: 20005-3934

15 (v) COMPUTER READABLE FORM:

16 (A) MEDIUM TYPE: Floppy disk

17 (B) COMPUTER: IBM PC compatible

18 (C) OPERATING SYSTEM: PC-DOS/MS-DOS

19 (D) SOFTWARE: PatentIn Release #1.0, Version #1.30

20 (vi) CURRENT APPLICATION DATA:

C--&gt; 21 (A) APPLICATION NUMBER: US/09/891,332A

C--&gt; 22 (B) FILING DATE: 27-Jun-2001

23 (C) CLASSIFICATION:

24 (vii) PRIOR APPLICATION DATA:

25 (A) APPLICATION NUMBER: 09/019,160

26 (B) FILING DATE:

27 (A) APPLICATION NUMBER: US 60/037,393

28 (B) FILING DATE: 07-FEB-1997

29 (viii) ATTORNEY/AGENT INFORMATION:

30 (A) NAME: Esmond, Robert W.

31 (B) REGISTRATION NUMBER: 32,893

32 (C) REFERENCE/DOCKET NUMBER: 0942.4250002

33 (ix) TELECOMMUNICATION INFORMATION:

34 (A) TELEPHONE: 202-371-2600

35 (B) TELEFAX: 202-371-2540

36 (2) INFORMATION FOR SEQ ID NO: 1:

37 (i) SEQUENCE CHARACTERISTICS:

38 (A) LENGTH: 2682 base pairs

39 (B) TYPE: nucleic acid

40 (C) STRANDEDNESS: both

41 (D) TOPOLOGY: both

42 (ii) MOLECULE TYPE: cDNA

ENTERED

## RAW SEQUENCE LISTING

DATE: 02/28/2002

PATENT APPLICATION: US/09/891,332A

TIME: 15:05:17

Input Set : N:\Crif3\Refhold\I891332A.raw

Output Set: N:\CRF3\02282002\I891332A.raw

```

43 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:
44 ATGGCGAGAC TATTTCTCTT TGATGGCACA GCCCTGGCCT ACAGGGCATA TTACGCCCTC 60
45 GACAGATCCC TTTCCACATC CACAGGAATT CCAACGAACG CCGTCTATGG CGTTGCCAGG 120
46 ATGCTCGTTA AATTCATTAA GGAACACATT ATACCCGAAA AGGACTACGC GGCTGTGGCC 180
47 TTCGACAAGA AGGCAGCGAC GTTCAGACAC AAAGTGTCTG TAAGCGACAA GGCGCAAAGG 240
48 CCAAAGACTC CGGCTCTTCT AGTTCAGCAC CTACCTTACA TCAAGCGGCT GATAGAAGCT 300
49 CTTGGTTTCA AAGTGCTGGA GCTGGAGGGA TACGAAGCAG ACGATATCAT CGCCACGCTT 360
50 GCAGTCAGGG CTGCACGTTT TTTGATGAGA TTTTCATTAA TAACCGGTGA CAAGGATATG 420
51 CTTCAACTTG TAAACGAGAA GATAAAGGTC TGGAGAATCG TCAAGGGGAT ATCGGATCTT 480
52 GAGCTTTACG ATTCGAAAAA GGTGAAAGAA AGATACGGTG TGGAAACCACA TCAGATACCG 540
53 GATCTTCTAG CACTGACGGG AGACGACATA GACAACATTC CCGGTGTAAC GGGAAATAGGT 600
54 GAAAAGACCG CTGTACAGCT TCTCGGCAAG TATAGAAATC TTGAATACAT TCTGGAGCAT 660
55 GCCCGTGAAC TCCCCCAGAG AGTGAGAAAAG GCTCTCTTGA GAGACAGGGA AGTTGCCATC 720
56 CTCAGTAAAA AACTTGCAAC TCTGGTGACG AACGCACCTG TTGAAGTGGA CTGGGAAGAG 780
57 ATGAAATACA GAGGATACGA CAAGAGAAAA CTACTTCCGA TATTGAAAGA ACTGGAGTTT 840
58 GCTTCCATCA TGAAGGAAC TCAACTGTAC GAAGAAGCAG AACCCACCGG ATACGAAATC 900
59 GTGAAGGATC ATAAGACCTT CGAAGATCTC ATCGAAAAGC TGAAGGAGGT TCCATCTTTT 960
60 GCCCTGGACC TTGAAACGTC CTCCCTTGAC CCGTTCAACT GTGAGATAGT CGGCATCTCC 1020
61 GTGTCGTTCA AACCGAAAAAC AGCTTATTAC ATTCCACTTC ATCACAGAAA CGCCCAGAAAT 1080
62 CTTGATGAAA CACTGGTGCT GTCGAAGTTG AAAGAGATCC TCGAAGACCC GTCTTCGAAG 1140
63 ATTGTGGGTC AGAACCTGAA GTACGACTAC AAGGTTCTTA TGGTAAAGGG TATATCGCCA 1200
64 GTTTATCCCG ATTTTGACAC GATGATAGCT GCATATTTGC TGGAGCCAAA CGAGAAAAAA 1260
65 TTCAATCTCG AAGATCTGTC TTTGAAATTT CTCGGATACA AAATGACGTC TTATCAGGAA 1320
66 CTGCTCGTCT TTTCCCTCACC ACTTTTGGT TTCAGCTTTG CGGATGTTCC GGTAGACAAG 1380
67 GCTGCGAACT ACTCCTGCGA GGATGCAGAC ATCACTTATA GGCTCTACAA GATACTCAGC 1440
68 ATGAAGCTCC ATGAAGCGGA ACTTGAGAAC GTCTTCTACA GGATAGAGAT GCCGTGGTG 1500
69 AACGTTCTTG CACGCATGGA ATTGAACGGG GTGTATGTGG ACACAGAATT CCTGAAAAAG 1560
70 CTCTCGGAGG AGTACGGCAA AAAGCTCGAG GAACTGGCCG AAAAAATCTA CCAGATAGCA 1620
71 GGTGAGCCCT TCAACATCAA TTCTCCAAAA CAGGTTTCAA AGATCCTTTT TGAGAAGCTG 1680
72 GGAATAAAAC CCCGTGGAAG AACGACAAAA ACAGGAGAGT ACTCTACCAG GATAGAGGTG 1740
73 TTGGAAGAGA TAGCGAATGA GCACGAGATA GTACCCCTCA TTCTCGAGTA CAGAAAGATC 1800
74 CAGAAACTGA AATCGACCTA CATAGACACC CTTCCGAAAC TTGTGAACCC GAAAACCGGA 1860
75 AGAATTATG CATCTTTCCA CCAGACGGGT ACCGCCACTG GCAGGTTGAG TAGCAGTGAT 1920
76 CCAAATCTTC AGAATCTTCC GACAAAGAGC GAAGAGGGAA AAGAAATTAG AAAAGCGATT 1980
77 GTGCCCCAGG ATCCAGACTG GTGGATCGTC AGTGCGGATT ATTCCCAAAT AGAATCAGA 2040
78 ATCCTCGCTC ATCTCAGTGG TGATGAGAAC CTTGTGAAGG CCTTCGAGGA GGGCATCGAT 2100
79 GTGCACACCT TGACTGCCTC CAGGATCTAC AACGTAAAGC CAGAAGAAGT GAACGAAGAA 2160
80 ATGCTACGGG TTGGAAGAT GGTGAACCTC TCTATAATAT ACGGTGTCAC ACCGTACGGT 2220
81 CTTTCTGTGA GACTTGGAAT ACCGGTTAAA GAAGCAGAAA AGATGATTAT CAGCTATTTT 2280
82 ACAGTGTATC CAAAGGTGCG AAGCTACATC CAGCAGGTTG TTGCAGAGGC AAAAGAGAAG 2340
83 GGCTACGTCA GGAATCTCTT TGGAAGAAAA AGAGATATTC CCCAGCTCAT GGCAAGGGAC 2400
84 AAGAACACCC AGTCCGAAGG CGAAAGAATC GCAATAAACA CCCCATTCA GGAACGGCG 2460
85 GCAGATATAA TAAATTTGGC TATGATAGAT ATAGACGAGG AGCTGAGAAA AAGAAACATG 2520
86 AAATCCAGAA TGATCATTCA GGTTCATGAC GAACTGGTCT TCGAGGTTCC CGATGAGGAA 2580
87 AAAGAAGAAC TAGTTGATCT GGTGAAGAAC AAAATGACAA ATGTGGTGAA ACTCTCTGTG 2640
88 CCTCTTGAGG TTGACATAAG CATCGAAAAA AGCTGGTCTT GA 2682
90 (2) INFORMATION FOR SEQ ID NO: 2:
91 (i) SEQUENCE CHARACTERISTICS:
92 (A) LENGTH: 893 amino acids

```

## RAW SEQUENCE LISTING

DATE: 02/28/2002

PATENT APPLICATION: US/09/891,332A

TIME: 15:05:17

Input Set : N:\Crf3\Refhold\I891332A.raw

Output Set: N:\CRF3\02282002\I891332A.raw

```

93      (B) TYPE: amino acid
94      (C) STRANDEDNESS: not relevant
W--> 95      (D) TOPOLOGY: not relevant
96      (ii) MOLECULE TYPE: protein
97      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:
98      Met Ala Arg Leu Phe Leu Phe Asp Gly Thr Ala Leu Ala Tyr Arg Ala
99      1          5          10          15
100     Tyr Tyr Ala Leu Asp Arg Ser Leu Ser Thr Ser Thr Gly Ile Pro Thr
101           20          25          30
102     Asn Ala Val Tyr Gly Val Ala Arg Met Leu Val Lys Phe Ile Lys Glu
103           35          40          45
104     His Ile Ile Pro Glu Lys Asp Tyr Ala Ala Val Ala Phe Asp Lys Lys
105           50          55          60
106     Ala Ala Thr Phe Arg His Lys Leu Leu Val Ser Asp Lys Ala Gln Arg
107           65          70          75          80
108     Pro Lys Thr Pro Ala Leu Leu Val Gln Gln Leu Pro Tyr Ile Lys Arg
109           85          90          95
110     Leu Ile Glu Ala Leu Gly Phe Lys Val Leu Glu Leu Glu Gly Tyr Glu
111           100         105         110
112     Ala Asp Asp Ile Ile Ala Thr Leu Ala Val Arg Ala Ala Arg Phe Leu
113           115         120         125
114     Met Arg Phe Ser Leu Ile Thr Gly Asp Lys Asp Met Leu Gln Leu Val
115           130         135         140
116     Asn Glu Lys Ile Lys Val Trp Arg Ile Val Lys Gly Ile Ser Asp Leu
117           145         150         155         160
118     Glu Leu Tyr Asp Ser Lys Lys Val Lys Glu Arg Tyr Gly Val Glu Pro
119           165         170         175
120     His Gln Ile Pro Asp Leu Leu Ala Leu Thr Gly Asp Asp Ile Asp Asn
121           180         185         190
122     Ile Pro Gly Val Thr Gly Ile Gly Glu Lys Thr Ala Val Gln Leu Leu
123           195         200         205
124     Gly Lys Tyr Arg Asn Leu Glu Tyr Ile Leu Glu His Ala Arg Glu Leu
125           210         215         220
126     Pro Gln Arg Val Arg Lys Ala Leu Leu Arg Asp Arg Glu Val Ala Ile
127           225         230         235         240
128     Leu Ser Lys Lys Leu Ala Thr Leu Val Thr Asn Ala Pro Val Glu Val
129           245         250         255
130     Asp Trp Glu Glu Met Lys Tyr Arg Gly Tyr Asp Lys Arg Lys Leu Leu
131           260         265         270
132     Pro Ile Leu Lys Glu Leu Glu Phe Ala Ser Ile Met Lys Glu Leu Gln
133           275         280         285
134     Leu Tyr Glu Glu Ala Glu Pro Thr Gly Tyr Glu Ile Val Lys Asp His
135           290         295         300
136     Lys Thr Phe Glu Asp Leu Ile Glu Lys Leu Lys Glu Val Pro Ser Phe
137           305         310         315         320
138     Ala Leu Asp Leu Glu Thr Ser Ser Leu Asp Pro Phe Asn Cys Glu Ile
139           325         330         335
140     Val Gly Ile Ser Val Ser Phe Lys Pro Lys Thr Ala Tyr Tyr Ile Pro
141           340         345         350

```

## RAW SEQUENCE LISTING

DATE: 02/28/2002

PATENT APPLICATION: US/09/891,332A

TIME: 15:05:17

Input Set : N:\Crf3\Refhold\I891332A.raw

Output Set: N:\CRF3\02282002\I891332A.raw

```

142   Leu His His Arg Asn Ala Gln Asn Leu Asp Glu Thr Leu Val Leu Ser
143           355                      360                      365
144   Lys Leu Lys Glu Ile Leu Glu Asp Pro Ser Ser Lys Ile Val Gly Gln
145           370                      375                      380
146   Asn Leu Lys Tyr Asp Tyr Lys Val Leu Met Val Lys Gly Ile Ser Pro
147   385                      390                      395                      400
148   Val Tyr Pro His Phe Asp Thr Met Ile Ala Ala Tyr Leu Leu Glu Pro
149           405                      410                      415
150   Asn Glu Lys Lys Phe Asn Leu Glu Asp Leu Ser Leu Lys Phe Leu Gly
151           420                      425                      430
152   Tyr Lys Met Thr Ser Tyr Gln Glu Leu Met Ser Phe Ser Ser Pro Leu
153           435                      440                      445
154   Phe Gly Phe Ser Phe Ala Asp Val Pro Val Asp Lys Ala Ala Asn Tyr
155           450                      455                      460
156   Ser Cys Glu Asp Ala Asp Ile Thr Tyr Arg Leu Tyr Lys Ile Leu Ser
157   465                      470                      475                      480
158   Met Lys Leu His Glu Ala Glu Leu Glu Asn Val Phe Tyr Arg Ile Glu
159           485                      490                      495
160   Met Pro Leu Val Asn Val Leu Ala Arg Met Glu Leu Asn Gly Val Tyr
161           500                      505                      510
162   Val Asp Thr Glu Phe Leu Lys Lys Leu Ser Glu Glu Tyr Gly Lys Lys
163           515                      520                      525
164   Leu Glu Glu Leu Ala Glu Lys Ile Tyr Gln Ile Ala Gly Glu Pro Phe
165           530                      535                      540
166   Asn Ile Asn Ser Pro Lys Gln Val Ser Lys Ile Leu Phe Glu Lys Leu
167   545                      550                      555                      560
168   Gly Ile Lys Pro Arg Gly Lys Thr Thr Lys Thr Gly Glu Tyr Ser Thr
169           565                      570                      575
170   Arg Ile Glu Val Leu Glu Glu Ile Ala Asn Glu His Glu Ile Val Pro
171           580                      585                      590
172   Leu Ile Leu Glu Tyr Arg Lys Ile Gln Lys Leu Lys Ser Thr Tyr Ile
173           595                      600                      605
174   Asp Thr Leu Pro Lys Leu Val Asn Pro Lys Thr Gly Arg Ile His Ala
175           610                      615                      620
176   Ser Phe His Gln Thr Gly Thr Ala Thr Gly Arg Leu Ser Ser Ser Asp
177   625                      630                      635                      640
178   Pro Asn Leu Gln Asn Leu Pro Thr Lys Ser Glu Glu Gly Lys Glu Ile
179           645                      650                      655
180   Arg Lys Ala Ile Val Pro Gln Asp Pro Asp Trp Trp Ile Val Ser Ala
181           660                      665                      670
182   Asp Tyr Ser Gln Ile Glu Leu Arg Ile Leu Ala His Leu Ser Gly Asp
183           675                      680                      685
184   Glu Asn Leu Val Lys Ala Phe Glu Glu Gly Ile Asp Val His Thr Leu
185           690                      695                      700
186   Thr Ala Ser Arg Ile Tyr Asn Val Lys Pro Glu Glu Val Asn Glu Glu
187   705                      710                      715                      720
188   Met Arg Arg Val Gly Lys Met Val Asn Phe Ser Ile Ile Tyr Gly Val
189           725                      730                      735
190   Thr Pro Tyr Gly Leu Ser Val Arg Leu Gly Ile Pro Val Lys Glu Ala

```

## RAW SEQUENCE LISTING

DATE: 02/28/2002

PATENT APPLICATION: US/09/891,332A

TIME: 15:05:17

Input Set : N:\Crf3\Refhold\I891332A.raw

Output Set: N:\CRF3\02282002\I891332A.raw

```

191          740          745          750
192  Glu Lys Met Ile Ile Ser Tyr Phe Thr Leu Tyr Pro Lys Val Arg Ser
193          755          760          765
194  Tyr Ile Gln Gln Val Val Ala Glu Ala Lys Glu Lys Gly Tyr Val Arg
195          770          775          780
196  Thr Leu Phe Gly Arg Lys Arg Asp Ile Pro Gln Leu Met Ala Arg Asp
197  785          790          795          800
198  Lys Asn Thr Gln Ser Glu Gly Glu Arg Ile Ala Ile Asn Thr Pro Ile
199          805          810          815
200  Gln Gly Thr Ala Ala Asp Ile Ile Lys Leu Ala Met Ile Asp Ile Asp
201          820          825          830
202  Glu Glu Leu Arg Lys Arg Asn Met Lys Ser Arg Met Ile Ile Gln Val
203          835          840          845
204  His Asp Glu Leu Val Phe Glu Val Pro Asp Glu Glu Lys Glu Glu Leu
205          850          855          860
206  Val Asp Leu Val Lys Asn Lys Met Thr Asn Val Val Lys Leu Ser Val
207  865          870          875          880
208  Pro Leu Glu Val Asp Ile Ser Ile Gly Lys Ser Trp Ser
209          885          890

```

211 (2) INFORMATION FOR SEQ ID NO: 3:

212 (i) SEQUENCE CHARACTERISTICS:

213 (A) LENGTH: 677 amino acids

214 (B) TYPE: amino acid

215 (C) STRANDEDNESS: not relevant

216 (D) TOPOLOGY: not relevant

W--&gt;

217 (ii) MOLECULE TYPE: protein

218 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

```

219  Met Ser Leu His Ala Arg Glu Leu Pro Gln Arg Val Arg Lys Ala Leu
220  1          5          10          15
221  Leu Arg Asp Arg Glu Val Ala Ile Leu Ser Lys Lys Leu Ala Thr Leu
222          20          25          30
223  Val Thr Asn Ala Pro Val Glu Val Asp Trp Glu Glu Met Lys Tyr Arg
224          35          40          45
225  Gly Tyr Asp Lys Arg Lys Leu Leu Pro Ile Leu Lys Glu Leu Glu Phe
226          50          55          60
227  Ala Ser Ile Met Lys Glu Leu Gln Leu Tyr Glu Glu Ala Glu Pro Thr
228  65          70          75          80
229  Gly Tyr Glu Ile Val Lys Asp His Lys Thr Phe Glu Asp Leu Ile Glu
230          85          90          95
231  Lys Leu Lys Glu Val Pro Ser Phe Ala Leu Ala Leu Glu Thr Ser Ser
232          100          105          110
233  Leu Asp Pro Phe Asn Cys Glu Ile Val Gly Ile Ser Val Ser Phe Lys
234          115          120          125
235  Pro Lys Thr Ala Tyr Tyr Ile Pro Leu His His Arg Asn Ala Gln Asn
236          130          135          140
237  Leu Asp Glu Thr Leu Val Leu Ser Lys Leu Lys Glu Ile Leu Glu Asp
238  145          150          155          160
239  Pro Ser Ser Lys Ile Val Gly Gln Asn Leu Lys Tyr Asp Tyr Lys Val
240          165          170          175

```

## VERIFICATION SUMMARY

PATENT APPLICATION: US/09/891,332A

DATE: 02/28/2002

TIME: 15:05:18

Input Set : N:\Crf3\Refhold\I891332A.raw

Output Set: N:\CRF3\02282002\I891332A.raw

L:21 M:220 C: Keyword misspelled or invalid format, [(A) APPLICATION NUMBER:]  
L:22 M:220 C: Keyword misspelled or invalid format, [(B) FILING DATE:]  
L:95 M:246 W: Invalid value of Alpha Sequence Header Field, [TOPOLOGY:], SeqNo=2  
L:216 M:246 W: Invalid value of Alpha Sequence Header Field, [TOPOLOGY:], SeqNo=3  
L:311 M:246 W: Invalid value of Alpha Sequence Header Field, [TOPOLOGY:], SeqNo=4  
L:398 M:246 W: Invalid value of Alpha Sequence Header Field, [TOPOLOGY:], SeqNo=5  
L:497 M:246 W: Invalid value of Alpha Sequence Header Field, [TOPOLOGY:], SeqNo=6  
L:618 M:246 W: Invalid value of Alpha Sequence Header Field, [TOPOLOGY:], SeqNo=7  
L:739 M:246 W: Invalid value of Alpha Sequence Header Field, [TOPOLOGY:], SeqNo=8  
L:860 M:246 W: Invalid value of Alpha Sequence Header Field, [TOPOLOGY:], SeqNo=9  
L:981 M:246 W: Invalid value of Alpha Sequence Header Field, [TOPOLOGY:], SeqNo=10  
L:1075 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:11